On the Consistency and Confidence of Distributed Dynamic State Estimation in Wireless Sensor Networks

Shaocheng Wang and Wei Ren

Abstract— **The problem of distributed dynamic state estimation in wireless sensor networks is studied. Two important properties of local estimates, namely, the consistency and confidence, are emphasized. On one hand, the consistency, which means that the approximated error covariance is lower bounded by the true unknown one, has to be guaranteed so that the estimate is not over-confident. On the other hand, since the confidence indicates the accuracy of the estimate, the estimate should be as confident as possible. We first analyze two different information fusion strategies used in the case of information sources with, respectively, uncorrelated errors and unknown but correlated errors. Then a distributed hybrid information fusion algorithm is proposed, where each agent uses the information obtained not only by itself, but also from its neighbors through communication. The proposed algorithm not only guarantees the consistency of the estimates, but also utilizes the available information sources in a more efficient manner and hence improves the confidence. Besides, the proposed algorithm is fully distributed and guarantees convergence with the sufficient condition formulated. The comparisons with existing algorithms are shown.**

I. INTRODUCTION

One of the most fundamental but important properties that an estimator should have is the *consistency* [1], [2]. That is, the approximated error covariance of an estimator should be lower bounded by the true error covariance. The approximated error covariance of an inconsistent estimate does not actually indicate its uncertainty [3]. This would cause issues in performance when it is further used in the downstream functions (data associations etc.). The consistency is especially important in the information fusion of distributed sensor networks. When fusing the information sources with unknown correlations, simply ignoring the unknown correlations might cause inconsistency in estimation. The *Covariance Intersection* (CI, [2]) algorithm is proposed to guarantee the consistency of estimates in the fusion process without knowing the cross-correlations between information sources. On the other hand, although an estimate should be consistent, it should not be too conservative. An over-conservative estimate indicates less confidence, which further indicates that the estimate is less useful. Therefore, it is significant to maximize the confidence while maintaining the consistency of the estimates.

Related work. Ref. [4] considers the consensus algorithm from the Kalman filters' perspective. The algorithm is further modified in [5] so that unbiased estimates can still be obtained even if the outflows of each agent are not equal to each other. Both algorithms in [4] and [5] are proposed to estimate a static state. Ref. [6] proposes the well-known Kalman-consensus filter (KCF), which combines the consensus filter and the local Kalman filter, and estimates the state of a linear dynamic process. The optimality and stability of the KCF is analyzed in [7]. The KCF requires the assumption that each agent and its neighbors have joint observability of the state of interest. If a state of interest is observed by neither a certain agent nor any of its local neighbors, this agent is referred as being *naive* about the state [8], [9]. An example of camera networks is shown in [8], where the generalized KCF (GKCF) is proposed, and is shown to outperform the KCF especially when there exist naive agents. The same situation is considered in [9], where the information-consensus filter (ICF) is proposed, and is shown to asymptotically approach the centralized Kalman filter (CKF) when there are infinite communication steps in between local updates. Recently, Ref. [10] proposes an innovative approach by achieving consensus on the local probability density functions (PDFs), referred as the "Kullback-Leibler average" (KLA) therein. It is analytically shown in [10] that each agent will eventually achieve bounded approximated estimate error covariance if the graph is strongly connected, even if each agent communicates with its neighbors for only once before updating its local estimate.

Contributions. While the current literature primarily focuses on the recovery of centralized estimators or the estimate errors of distributed algorithms, in this paper, we emphasize on the importance of the consistency and confidence of estimates. Considered from the perspective of information fusion, a *distributed hybrid information fusion* (DHIF) algorithm is proposed. In the proposed DHIF, each agent cooperates with its neighbors, utilizes all available information sources, and comes up with its local estimate that is consistent while relatively confident. We focus on a more realistic scenario where only one communication step is allowed between neighbors before they update their local estimates. The proposed algorithm is robust against naive agents and does not require any global parameters. The sufficient condition for bounded approximated local estimate error covariance is formulated. The comparisons with existing algorithms are shown.

II. PRELIMINARIES

A. Notations

Through this paper, we use the following notations. $\mathbf{1}$ is an all-one vector with appropriate dimension. $\mathbf{0}$ is an all-zero matrix with appropriate dimension. $\mathbf{I}_n$ is the $n \times n$ identity

Shaocheng Wang and Wei Ren are with the University of California, Riverside, CA 92507, USA. Email: shaocheng.wang@email.ucr.edu, ren@ece.ucr.edu. This paper has been accepted by the 54th IEEE Conference on Decision and Control in December, 2015 at Osaka, Japan.

matrix. $\mathbb{E}[\cdot]$ denotes the expectation of a random variable. $\mathbb{Z}^*$ is the set of non-negative integers. $\mathbb{Z}^+$ is the set of positive integers. For any arbitrary square matrices X_1 and X_2 with the same dimensions, $X_1 \succeq X_2$ (respectively, $X_1 \succ X_2$) implies that $X_1 - X_2$ is a positive semi-definite (respectively, positive definite) matrix. $X_1 \geq X_2$ (respectively, $X_1 > X_2$) implies that $X_1 - X_2$ is a non-negative (respectively, positive) matrix. For an arbitrary matrix X, $\{X\}_{(i,j)}$ is the entry on its ith row, jth column. $\mathbf{tr}(X)$ is the trace of X. When we refer $X \succ \mathbf{0}$ as a large/small matrix, we mean X is large/small in the sense of the mean squared error (MSE). If X is nonsingular, $X^{-n\top} \triangleq ((X^n)^{-1})^\top$ for some $n \in \mathbb{Z}^*$.

B. Graph Theory

A directed graph $\mathscr{G}(\mathscr{V}, \mathscr{E})$ is used to represent the communication topology of a large-scale sensor network, where $\mathscr{V} \triangleq \{1, \cdots, N\}$ and $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$ are, respectively, the set of N vertices that stands for the local agents, and the set of edges that stands for the communication channels. Several basic concepts of graph theory used later in this paper are briefly listed here. An *edge* $(i,j) \in \mathscr{E}$ denotes that agent j can receive information from agent i. If a graph is *undirected*, $(i,j) \in \mathscr{E}$ implies that $(j,i) \in \mathscr{E}$. A *directed path* from agent i_0 to agent i_ℓ is a sequence of vertices $i_0, i_1, \cdots, i_\ell$ such that $(i_{j-1}, i_j) \in \mathscr{E}$ for $0 < j \leq \ell$. A directed (respectively, undirected) graph is *strongly connected* (respectively, *connected*) if there exists at least one path from every vertex to every other vertex. $N_i \triangleq \{j | (j,i) \in \mathscr{E}, \forall j \neq i\}$ is the *neighborhood* of agent i. $J_i \triangleq N_i \cup \{i\}$ is the *inclusive neighborhood* of agent i. A graph is *complete* if $j \in N_i$ for any i and j. The *in-degree* of a certain agent i is defined as $\Delta_i \triangleq |N_i|$. The *maximum in-degree* of a graph is defined as $\Delta_{\max} \triangleq \max_{i \in \{1,\dots,N\}} \Delta_i$. A *directed spanning tree* is a subgraph of $\mathscr{G}(\mathscr{V}, \mathscr{E})$ such that this subgraph is a directed tree and contains all vertices of $\mathscr{G}(\mathscr{V}, \mathscr{E})$. For a graph $\mathscr{G}(\mathscr{V}, \mathscr{E})$, an associated *stochastic matrix* D is defined as follows. For each $i \in \mathscr{V}$, $\{D\}_{(i,j)} > 0$ if $j \in J_i$ and $\{D\}_{(i,j)} = 0$ otherwise. Moreover, $\sum_j \{D\}_{(i,j)} = 1$. If it also holds that $\sum_i \{D\}_{(i,j)} = 1$, $\forall j$, then D is a *doubly stochastic matrix*.

Definition 1 (Induced family of stochastic matrices): Let D be a stochastic matrix associated with the graph $\mathscr{G}(\mathscr{V}, \mathscr{E})$. Then the family of stochastic matrices induced by D is the set of all possible stochastic matrices associated with $\mathscr{G}(\mathscr{V}, \mathscr{E})$.

C. Discrete-time Average Consensus Algorithm

An average consensus algorithm computes the global average of some variables of interest, through communication with only local neighbors. Suppose that $\mathbf{a}_i[0]$ is the initial value of agent i's local variable. The objective is to compute $\frac{1}{N} \sum_{i=1}^N \mathbf{a}_i[0]$ in a distributed manner. Suppose that the graph is undirected, the local algorithm implemented at agent i has the form of $\mathbf{a}_i[k+1] = \mathbf{a}_i[k] + \varepsilon \sum_{j \in N_i} (\mathbf{a}_j[k] - \mathbf{a}_i[k])$, $\forall k \in \mathbb{Z}^*$, where $\varepsilon \in (0, 1/\Delta_{\max})$ is the rate parameter. By iteratively updating the local variable value in this manner, the global average can be asymptotically achieved if and only if the graph is connected. More details about the average consensus algorithm can be found in [11].

D. Linear Information Fusion

Consider a set of p pieces of information sources, denoted as $\{a_i\}_{i=1}^p$, where each $a_i \in \mathbb{R}^{m_i}$ is the addition of a quantity linearly related to the same unknown parameter of interest $\alpha \in \mathbb{R}^n$, and some unknown noise/errors with zero mean. That is, $\mathbb{E}[a_i] = C_i\alpha$ with $C_i \in \mathbb{R}^{m_i \times n}$, $\forall i$. For each a_i, assume that the noise/error covariance, denoted as $\tilde{R}_{a_i} \triangleq \mathbb{E}[(a_i - \mathbb{E}[a_i])(a_i - \mathbb{E}[a_i])^\top]$, is positive definite and might be unknown, but a *consistent* ([1], [2]) approximation, denoted as R_{a_i}, is known, i.e., $R_{a_i} \succeq \tilde{R}_{a_i} \succ \mathbf{0}$.

The objective is to linearly fuse $\{a_i\}_{i=1}^p$ and obtain an unbiased consistent estimate of α, i.e., $\hat{\alpha} = \sum_{i=1}^p K_i a_i$, where $K_i \in \mathbb{R}^{n \times m_i}$, $\forall i$ are weighting matrices to be selected. Note that $\sum_{i=1}^p K_i C_i = \mathbf{I}_n$ due to the required unbiasedness of $\hat{\alpha}$. Let $\tilde{R}_{\hat{\alpha}} \triangleq \mathbb{E}[(\hat{\alpha} - \mathbb{E}[\hat{\alpha}])(\hat{\alpha} - \mathbb{E}[\hat{\alpha}])^\top]$ be the true error covariance of $\hat{\alpha}$. Let $\tilde{R}_{a_{ij}} \triangleq \mathbb{E}[(a_i - \mathbb{E}[a_i])(a_j - \mathbb{E}[a_j])^\top]$. It follows that

$$\tilde{R}_{\hat{\alpha}} = \sum_i K_i \tilde{R}_{a_i} K_i^\top + \sum_i \sum_{j \neq i} K_i \tilde{R}_{a_{ij}} K_j^\top. \quad (1)$$

When $\tilde{R}_{a_i}$ and $\tilde{R}_{a_{ij}}$ are unknown, an approximation of $\tilde{R}_{\hat{\alpha}}$, is required. This is when the problem becomes tricky. On one hand, the consistency of the estimate has to be guaranteed since an inconsistent estimate is over-confident and the corresponding approximated error covariance does not realistically implies the uncertainty [3]. On the other hand, as the error covariance is a measure of estimate uncertainty. For a consistent estimate, a smaller approximated error covariance indicates more confidence. In the rest of this paper, we assume that when fusing information sources, the objective is to minimize the trace of the fused error covariance. Two situations are considered.

1) Fusion of information with uncorrelated errors: When the information sources have mutually uncorrelated errors, i.e., $\tilde{R}_{a_{ij}} = \mathbf{0}$, $\forall i,j$, it follows from (1) that $\tilde{R}_{\hat{\alpha}} = \sum_i K_i \tilde{R}_{a_i} K_i^\top$. Thus a reasonable approximation of $\tilde{R}_{\hat{\alpha}}$ could be $R_{\mathrm{u}} \triangleq \sum_i K_i R_{a_i} K_i^\top$. Note that if $R_{a_i} \succeq \tilde{R}_{a_i}$, $\forall i$, it is guaranteed that $R_{\mathrm{u}} \succeq \tilde{R}_{\hat{\alpha}}$. The optimal R_{u} and $\hat{\alpha}$, denoted as, respectively, R_{u}^* and $\hat{\alpha}^*$, can be obtained by minimizing $\mathbf{tr}(R_{\mathrm{u}})$ with $\sum_i K_i C_i = \mathbf{I}_n$ being held to guarantee the unbiasedness.

Lemma 1: Let K_i^* be the optimal weighting matrices such that $\mathbf{tr}(\sum_i K_i R_{a_i} K_i^\top)$ is minimized with $\sum_i K_i C_i = \mathbf{I}_n$ satisfied. Let $\hat{\alpha}^* \triangleq \sum_i K_i^* a_i$ and $R_{\mathrm{u}}^* \triangleq K_i^* R_{a_i} K_i^{*\top}$. Then, $K_i^* = (C_i^\top R_{a_i}^{-1} C_i)^{-1} C_i^\top R_{a_i}^{-1}$, $\forall i$ and

$$R_{\mathrm{u}}^* = (\sum_i C_i^\top R_{a_i}^{-1} C_i)^{-1}, \quad \hat{\alpha}^* = R_{\mathrm{u}}^* (\sum_i C_i^\top R_{a_i}^{-1} a_i). \quad (2)$$

Proof: Let $\Lambda \in \mathbb{R}^{n \times n}$ be the Lagrangian multiplier. Construct the following augmented function

$$L(K_1, \cdots, K_p, \Lambda) \triangleq \mathbf{tr} \sum_i K_i R_{a_i} K_i^\top + \sum_{j,k} \{\Lambda(\sum_i K_i C_i - \mathbf{I}_n)\}_{(j,k)}.$$

By setting the partial derivative with respect to Λ and each K_i to zero, it follows that $K_i^* = \frac{1}{2}\Lambda^* C_i^\top R_{a_i}^{-1}$, $\forall i$ and $\Lambda^* = 2(\sum_i C_i^\top R_{a_i}^{-1} C_i)^{-1}$. Accordingly, $\hat{\alpha}^*$ and R_{u}^* can be obtained by their definitions. ∎

In summary, the fusion steps given by (2) preserve the consistency, and yield the optimal estimate. Note that if $R_{a_i} = \tilde{R}_{a_i}, \forall i$, one can obtain $R_u^* = \tilde{R}_{\hat{\alpha}}$. However, this only holds when the information sources have mutually uncorrelated errors.

2) Fusion of information with unknown correlated errors: When the information sources have unknown but correlated errors, the last term in (1) is nonzero and unknown. In such a case, if one approximates $\tilde{R}_{\hat{\alpha}}$ with R_u as in the previous case, it is not guaranteed that $R_u \succeq \tilde{R}_{\hat{\alpha}}$ even if $R_{a_i} \succeq \tilde{R}_{a_i}, \forall i$. The Covariance Intersection (CI) algorithm is proposed in [2], where the estimate and approximated error covariance of α, denoted as, respectively, $\hat{\alpha}_c$ and R_c, are as follows:

$$R_c = (\sum_i \omega_i C_i^\top R_{a_i}^{-1} C_i)^{-1}, \quad \hat{\alpha}_c = R_c(\sum_i \omega_i C_i^\top R_{a_i}^{-1} a_i), \quad (3)$$

where $\omega_i > 0, \forall i$ are the weights satisfy $\sum_i \omega_i = 1$. The CI algorithm is able to preserve the consistency of the information sources (see [2] for detailed proofs), for any possible unknown correlated errors. However, the optimality is traded in exchange for guaranteeing the consistency. As a result, the CI algorithm could be too conservative especially when the information sources are less correlated.

III. PROBLEM FORMULATION

Through this paper, we use the sub-index $k \in \mathbb{Z}^+$ to denote each variable at the kth time instant. Consider the following linear time-invariant (LTI) dynamic system

$$x_{k+1} = Fx_k + Bw_k, \quad (4)$$

where $x_k \in \mathbb{R}^n$ is the state of interest. $F \in \mathbb{R}^{n \times n}$ is the state transition matrix. $B \in \mathbb{R}^{n \times p}$ is a full rank matrix that models the process noise $w_k \in \mathbb{R}^p$, which is assumed to be white Gaussian, i.e., $w_k \sim \mathcal{N}(0, Q)$. Here $Q \in \mathbb{R}^{p \times p}$ is the covariance of the process noise and it is assumed that $Q \succ 0$. Moreover, it is assumed that $\mathbb{E}[x_{k'} w_k^\top] = 0, \forall k' \in \mathbb{Z}^*$ and $k > k'$.

Suppose that each agent $i, i \in \mathcal{V}$, is able to obtain a local measurement $z_{i,k} \in \mathbb{R}^{m_i}$, which follows a LTI sensing model:

$$z_{i,k} = H_i x_k + v_{i,k}, \quad (5)$$

where $H_i \in \mathbb{R}^{m_i \times n}$ is the local observation matrix. $v_{i,k}$ is the local measurement noise and assumed to be white Gaussian, i.e., $v_{i,k} \sim \mathcal{N}(0, R_i)$ with $R_i \in \mathbb{R}^{m_i \times m_i}$ and $R_i \succ 0$ being the covariance of the measurement noise.

In large-scale sensor networks, it is common to encounter situations where the target of interest is not directly observed by a subset of local agents. The following assumption is made through this paper.

Assumption 1: If the target of interest is not observed by agent i at time instant k, then $R_i^{-1} = 0$ and $z_{i,k}$ is arbitrary.

Assumption 1 essentially states that, an agent observing no target has infinite uncertainty about its measurement. As $z_{i,k}$ can be arbitrary for such an agent, one can regard either H_i or $v_{i,k}$ as arbitrary. From this point, to simplify the statement of the observability conditions, it is assumed that $H_i = 0$ if

the target is not observed by agent i at time k. The naive agent can be therefore defined as follows.

Definition 2 (Naive agent): Let $\mathfrak{N} \subset \mathcal{V}$ be the set of vertices corresponding to all naive agents in the network. Then $i \in \mathfrak{N}$ if $(F, \text{col}\{H_j\}_{j \in J_i})$ is not an observable pair.

In other words, an agent is naive as long as there exists one state component which cannot be recovered from the collective measurements in its inclusive neighborhood. In such a case, the agent is actually only naive with respect to this state component. In the rest of this paper, for simplicity, we focus on naive agents that can recover none of the state components. In the case where naivety is discussed with respect to some state components, the same conclusions hold for the corresponding subspaces of the state space.

The objective is formulated as follows. Each agent is equipped with a local sensor formulated by (5). The target makes uses of both the local information sources, and the information obtained from its neighbors through communication to come up with a consistent estimate of the dynamic state in (4), with as much confidence as possible. The estimation process should not depend on any global information.

IV. DISTRIBUTED ESTIMATION WITH HYBRID INFORMATION FUSION STRATEGY

Suppose that at time instant k, each agent is able to take its local measurement $z_{i,k}$. Besides $z_{i,k}$, each agent has a consistent prior estimate of the state of interest, denoted as $\hat{x}_{i,k|k-1}$. The local prior estimate error and approximated error covariance are defined as, respectively, $\hat{\eta}_{i,k|k-1} \triangleq \hat{x}_{i,k|k-1} - x_k$ and $P_{i,k|k-1} \succeq \tilde{P}_{i,k|k-1}$, where $\tilde{P}_{i,k|k-1} \triangleq \mathbb{E}[\hat{\eta}_{i,k|k-1} \hat{\eta}_{i,k|k-1}^\top]$ is the true error covariance of $\hat{x}_{i,k|k-1}$. The following assumptions are made through this paper.

Assumption 2: For each $k \in \mathbb{Z}^+$: (2.1) All local measurement noises are mutually uncorrelated, i.e., $\mathbb{E}[v_{i,k} v_{j,k}^\top] = 0, \forall i \neq j$; (2.2) Any local prior estimate error is uncorrelated with any local measurement noise, $\mathbb{E}[\hat{\eta}_{i,k|k-1} v_{j,k}^\top] = 0, \forall i, j$.

A. Distributed Hybrid Information Fusion (DHIF)

Now we discuss the strategy to fuse all information sources available to agent i, at $k \in \mathbb{Z}^+$. The proposed strategy is a hybrid of two fusion steps.

Step 1. Fuse the prior estimates obtained from all agents in the inclusive neighborhood, i.e., $\{\hat{x}_{j,k|k-1}\}_{j \in J_i}$, and obtain an intermediate estimate and its approximated error covariance, denoted as, respectively, $\check{x}_{i,k}$ and $\check{P}_{i,k}$. Since the correlations between each pair of the prior estimate errors in the inclusive neighborhood are neither known nor negligible, the CI algorithm [2] is used for this step. Let I_n, $P_{j,k|k-1}$, $\hat{x}_{j,k|k-1}, \forall j \in J_i$, $\check{P}_{i,k}$ and $\check{x}_{i,k}$ play, respectively, the roles of $C_i, R_{a_i}, a_i, \forall i \in \{1, \cdots, p\}$, R_c and $\hat{\alpha}_c$ in (3). It follows that $\check{P}_{i,k}^{-1} = \sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1}$, $\check{P}_{i,k}^{-1} \check{x}_{i,k} = \sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1} \hat{x}_{j,k|k-1}$, where $\omega_{ij}^{(k)} \in \mathbb{R}^+$ is the weight that agent i assigns to the information received from agent j, at time instant k. The selection of $\omega_{ij}^{(k)}$ will be discussed in detail in Section IV-B. Note that the CI algorithm requires that $\sum_{j \in J_i} \omega_{ij}^{(k)} = 1$ for any given i or k.

Step 2. Fuse the intermediate estimate obtained from Step 1 (i.e., $\check{x}_{i,k}$) with all measurements obtained from the inclusive neighborhood (i.e., $\{z_j\}_{j \in J_i}$), and obtain the posterior state estimates and its approximated error covariance. The conclusion of the following lemma is necessary for this step.

Lemma 2: Under Assumptions (2.1) and (2.2), $\{v_{j,k}\}_{j \in J_i}$ and $(\check{x}_{i,k} - x_k)$ are mutually uncorrelated, $\forall i \in \mathscr{V}$.

Proof: With Assumption (2.1) satisfied, it suffices to show that for each $l \in J_i$, $\mathbb{E}[(\check{x}_{i,k} - x_k)v_{l,k}^\top] = \mathbf{0}$. This can be verified as

$$\mathbb{E}[(\check{x}_{i,k} - x_k)v_{l,k}^\top] = \mathbb{E}[(\check{P}_{i,k} \sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1} \hat{x}_{j,k|k-1}$$

$$-\check{P}_{i,k}\check{P}_{i,k}^{-1} x_k)v_{l,k}^\top] = \check{P}_{i,k} \sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1} \mathbb{E}[\hat{\eta}_{j,k|k-1}v_{l,k}^\top] = \mathbf{0},$$

where Assumption (2.2) is used for the last equality. ∎

As shown by Lemma 2, all $|J_i| + 1$ information sources to fuse have mutually uncorrelated errors. Therefore, Eq. (2) can be applied. Without loss of generality, let H_j, R_j and $z_{j,k}$, $\forall j \in J_i$ play, respectively, the roles of C_i, R_{a_i} and a_i, $\forall i \in \{1, \cdots, p-1\}$ in (2), let $\mathbf{I}_n$, $\check{P}_{i,k}$, $\check{x}_{i,k}$, $P_{j,k|k}$ and $\hat{x}_{j,k|k}$ play, respectively, the roles of C_p, R_{a_p}, a_p, R_u^* and $\hat{\alpha}^*$ in (2), with $\check{P}_{i,k}$ and $\check{x}_{i,k}$ obtained from the previous step, it follows that

$$P_{i,k|k} = (\sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1} + \sum_{j \in J_i} H_j^\top R_j^{-1} H_j)^{-1},$$

$$\hat{x}_{i,k|k} = P_{i,k|k}(\sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1} \hat{x}_{j,k|k-1} + \sum_{j \in J_i} H_j^\top R_j^{-1} z_j). \quad (6)$$

We refer (6) as the update steps of the proposed algorithm, named Distributed Hybrid Information Fusion (DHIF), implemented by agent i at time instant k. Here $\hat{x}_{i,k|k}$ and $P_{i,k|k}$ in (6) are, respectively, the local posterior estimate and the approximated local posterior error covariance by the DHIF. Define the posterior estimate error and the true posterior error covariance as, respectively, $\hat{\eta}_{i,k|k} \triangleq \hat{x}_{i,k|k} - x_k$ and $\tilde{P}_{i,k|k} \triangleq \mathbb{E}[\hat{\eta}_{i,k|k}\hat{\eta}_{i,k|k}^\top]$. The consistency is preserved in this 2-step fusion process, as stated in the following lemma.

Lemma 3: Under Assumptions (2.1) and (2.2), the update steps (6) preserve the consistency, i.e., $P_{i,k|k} \succeq \tilde{P}_{i,k|k}$ if $\{z_{j,k}\}_{j \in J_i}$ and $\{\hat{x}_{i,k|k-1}\}_{j \in J_i}$ are all consistent information sources.

Proof: The fact that $\hat{x}_{i,k|k}$ given by (6) preserves consistency can be directly observed because both Step 1 and Step 2 are fusion strategies that preserve the consistency. ∎

B. Weights Selection for CI Algorithm

Because the CI algorithm is used in Step 1, the update steps in (6) only give a suboptimal estimate. As naive agents could come up with estimates with very low confidence, if relatively high weights are assigned to such estimates, the intermediate estimate obtained by Step 1 will become less confident. Therefore, the selection of the weights for the embedded CI algorithm in Step 1 should be determined carefully. This is especially important in the scenario where the agents are allowed to communicate only once with their neighbors in between, as the advantages of asymptotic

properties brought by the consensus-based algorithms are lost in such a scenario. The optimal set of $\{\omega_{ij}^{(k)}\}_{j \in J_i}$ can be obtained by minimizing $\mathbf{tr}(\sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1})^{-1}$, which can be casted as the following Semi-definite Programming (SDP) problem [12]:

$$\underset{u}{\text{minimize}} \ u^\top \mathbf{1},$$

$$\text{subject to} \sum_{j \in J_i} \omega_{ij}^{(k)} = 1, \ 0 < \omega_i \le \omega_{ij}^{(k)} \le 1, \ \forall j \in J_i,$$

$$\begin{bmatrix} \sum_{j \in J_i} \omega_{ij}^{(k)} P_{j,k|k-1}^{-1} & e_l \\ e_l^\top & u_{(l)} \end{bmatrix} \succeq \mathbf{0}, \ l = 1, \cdots, n, \quad (7)$$

where $\underline{\omega}_i$, $\forall i$ are sufficiently small constant lower bounds for selecting weights. Here $e_l \in \mathbb{R}^n$ is the canonical basis vector whose lth entry is one and $u \in \mathbb{R}^n$. Due to the convexity of the problem formulated in (7), the set of weights can be determined efficiently. In the case where solving (7) is still considered as computationally expensive, some suboptimal approximations can still be used [13].

C. Recursive Form

Algorithm 1: DHIF Implemented by Agent i at Time k

---

1 **if** $k = 1$ **then**
2 $\quad$ initializes $\hat{x}_{i,1|0}$ and $P_{i,1|0}$
3 computes $\Xi_{i,k} \triangleq P_{i,k|k-1}^{-1}$ and $\xi_{i,k} \triangleq P_{i,k|k-1}^{-1} \hat{x}_{i,k|k-1}$
4 takes the local measurement $z_{i,k}$
5 computes $S_{i,k} \triangleq H_i^\top R_i^{-1} H_i$ and $y_{i,k} \triangleq H_i^\top R_i^{-1} z_{i,k}$
6 sends $S_{i,k}$, $y_{i,k}$, $\Xi_{i,k}$ and $\xi_{i,k}$ to agent j, $\forall j$ such that $i \in N_j$
7 receives $S_{j,k}$, $y_{j,k}$, $\Xi_{j,k}$ and $\xi_{j,k}$ from agent j, $\forall j \in N_i$
8 selects the set of weights $\{\omega_{ij}^{(k)}\}_{j \in J_i}$
9 computes $\bar{\Xi}_{i,k} = \sum_{j \in J_i} \omega_{ij}^{(k)} \Xi_{i,k}$ and $\bar{\xi}_{i,k} = \sum_{j \in J_i} \omega_{ij}^{(k)} \xi_{i,k}$
10 computes $\bar{y}_{i,k} = \sum_{j \in J_i} y_{j,k}$ and $\bar{S}_{i,k} = \sum_{j \in J_i} S_{j,k}$
11 updates local estimate and approximated covariance

$$P_{i,k|k} = (\bar{S}_{i,k} + \bar{\Xi}_{i,k})^{-1} \quad (8)$$

$$\hat{x}_{i,k|k} = P_{i,k|k}(\bar{y}_{i,k} + \bar{\xi}_{i,k}) \quad (9)$$

12 predicts local estimate and approximated covariance

$$P_{i,k+1|k} = FP_{i,k|k}F^\top + BQB^\top \quad (10)$$

$$\hat{x}_{i,k+1|k} = F\hat{x}_{i,k|k} \quad (11)$$

With the prediction steps formulated based on the system dynamics in (4), the recursive form of the distributed hybrid information fusion (DHIF) strategy implemented by agent i at time k is summarized in Algorithm 1.

Theorem 1: Under Assumption 2, Algorithm 1 preserves the consistency. That is, $P_{i,k|k-1} \succeq \tilde{P}_{i,k|k-1}$ and $P_{i,k|k} \succeq \tilde{P}_{i,k|k}$, $\forall k \in \mathbb{Z}^+$, if the initialized local estimates $\{\hat{x}_{i,1|0}\}_{i=1}^N$ are consistent. That is,

$$P_{i,1|0} \succeq \mathbb{E}[(\hat{x}_{i,1|0} - x_1)(\hat{x}_{i,1|0} - x_1)^\top], \ \forall i. \quad (12)$$

Proof: Lemma 3 has shown that the update steps in (6) preserve the consistency of the information. Therefore, it suffices to show if $P_{i,k|k} \succeq \tilde{P}_{i,k|k}$, one can obtain $P_{i,k+1|k} \succeq \tilde{P}_{i,k+1|k}$. Note that $\tilde{P}_{i,k+1|k} = F\tilde{P}_{i,k|k}F^\top + BQB^\top + \Omega_1 + \Omega_2$, where $\Omega_1 = F\mathbb{E}[\hat{\eta}_{i,k|k}w_{k+1}^\top]B^\top$ and $\Omega_2 = \Omega_1^\top$. Since $\hat{\eta}_{i,k|k}$ in a linear combination of $\{x_0, \{w_l\}_{l=1}^k, \{v_l\}_{l=1}^k\}$, each of which is assumed to be uncorrelated with w_{k+1}, it follows that $\mathbb{E}[\hat{\eta}_{i,k|k}w_{k+1}^\top] = \mathbf{0}$. Therefore, $\Omega_2 = \Omega_1^\top = \mathbf{0}$. It follows from (8) that if $P_{i,k|k} \succeq \tilde{P}_{i,k|k}$, $P_{i,k+1|k} = FP_{i,k|k}F^\top + BQB^\top \succeq F\tilde{P}_{i,k|k}F^\top + BQB^\top = \tilde{P}_{i,k+1|k}$. Therefore, the prediction steps in Algorithm 1 also preserve the consistency. It follows that any estimate at any agent obtained by the recursive Algorithm 1 is consistent, if all local estimates initially fed to the algorithm are consistent, as formulated in (12). ∎

Remark 1: It is worth mentioning that (12) can be easily satisfied in general. The prior knowledge about the state of interest can be learned in an off-line manner before the fusion process. In the worst case, each agent can simply choose $P_{i,1|0}^{-1} = \mathbf{0}$, which indicates the infinite initial local uncertainty so that (12) is satisfied.

V. Comparisons with Existing Algorithms

In this section, we analyze the advantages of the proposed DHIF algorithm, and compare it with some existing algorithms in the literature.

A. Robustness in Presence of Naive Agents

Note that (9) can be written as

$$\hat{x}_{i,k|k} = \hat{x}_{i,k|k-1} + P_{i,k|k}\left(\bar{y}_{i,k} - \bar{S}_{i,k}\hat{x}_{i,k|k-1}\right) + P_{i,k|k}\sum_{j\in J_i}\omega_{ij}^{(k)}P_{j,k|k-1}^{-1}(\hat{x}_{j,k|k-1} - \hat{x}_{i,k|k-1}). \quad (13)$$

As observed in (13), the approximated posterior estimate error covariance $P_{i,k|k}$, which implies the uncertainty of agent i's local estimate, is multiplied to both of the last two terms. On one hand, if agent i is more confident about its own local estimate, $P_{i,k|k}$ will be small. In such a situation, the effects on $\hat{x}_{i,k|k}$ caused by the last two terms in (13) will be attenuated by a small $P_{i,k|k}$. On the other hand, when $P_{i,k|k}$ is large, (13) will push the local posterior estimate to match its neighbors' measurements and their local prior estimates, as described from the following two perspectives.

(1) Note that $\bar{y}_{i,k} - \bar{S}_{i,k}\hat{x}_{i,k|k-1} = \sum_{j\in J_i}H_j^\top R_j^{-1}(z_{i,k} - H_j\hat{x}_{i,k|k-1})$, which can be regarded as a weighted sum of the innovation terms that push $\hat{x}_{i,k|k}$ to match $z_{j,k}$, $\forall j \in J_i$. If a neighboring agent j does not directly observe the state of interest, $R_j^{-1} = \mathbf{0}$. Thus, its bad effects on the update of $\hat{x}_{i,k|k}$, caused by the erroneous local measurement $z_{j,k}$, will be eliminated in an automated manner.

(2) Similarly, the last term in (13) can also be regarded as a weighted sum, which pushes $\hat{x}_{i,k|k}$ to match $\hat{x}_{j,k|k-1}$, $\forall j \in J_i$. The weight that agent i assigns to agent j is the multiplication of $\omega_{ij}^{(k)}$ and $P_{j,k|k-1}^{-1}$. The former component is selected to increase the confidence of the estimate while preserving its consistency; the latter one, which implies the confidence of each neighbor's local prior estimate, pushes the local estimate toward its neighbors' with higher confidence.

Let δ be a scalar parameter. If the last term in (13) is replaced with $\delta P_{i,k|k}\sum_{j\in J_i}(\hat{x}_{j,k|k-1} - \hat{x}_{i,k|k-1})$, the update step of the local posterior estimates proposed by the KCF is obtained. The KCF [7] has been shown to perform well with guaranteed convergence under certain conditions. However, it should be noted that the KCF relies on the assumption that the state of interest is jointly observable in the inclusive neighborhood of every agent. The divergence of the KCF in presence of naive agents has been shown in [8] and [9]. Essentially with the last term of (13) replaced with $\delta P_{i,k|k}\sum_{j\in J_i}(\hat{x}_{j,k|k-1} - \hat{x}_{i,k|k-1})$ in the KCF, each agent equally weighs its neighbors' prior estimates regardless of the confidence of these estimates. The performance is therefore deteriorated especially when there exist some other naive agents in its neighborhood. More detailed analysis on the performance of the KCF in presence of naive agents can be found in [8].

B. Fully Distributed

In the GKCF [8] and ICF [9], the rate parameter ε is required to be chosen between 0 and $1/\Delta_{\max}$ to guarantee the performance of the embedded average consensus algorithm. If the maximum in-degree is changing with time, a proper selection of ε might not be as good as before. Even if $\Delta_{\max}$ is known, it is not clear how to select a nice ε in the studies of the GKCF and ICF. In [8], [9], after obtaining a new local measurement, each agent is allowed to communicate with its local neighbors for infinite times, before it finally updates its posterior local estimate at the current time instant. Therefore, in such a case, any selection between 0 and $1/\Delta_{\max}$ will guarantee that every agent's local estimate asymptotically becomes identical before its next measurement update. Unfortunately, in the realistic case where every agent communicates only once with its neighbors before updating its local estimate, the importance of selecting ε is similar to the selection of weights in the proposed DHIF, as discussed in Section IV-B.

Besides $\Delta_{\max}$, the ICF also requires each agent to know the total number of agents N in the network to asymptotically approach the centralized solution via infinite communication steps before the agents update their local estimates. Similar to the discussion on $\Delta_{\max}$, N could be changing over time. Moreover, in the case focused in this paper (single communication step), the ICF is not able to obtain the optimal solution even with the correct knowledge of N.

The proposed DHIF algorithm does not require any global information. It is run in an automated manner and is adaptive to the locally unknown changes in the network.

C. Consistency of Estimates

The consistency is one of the most fundamental but significant properties to be preserved during the information fusion process. That is, the approximated error covariance should be lower bounded by the true error covariance. The approximated error covariance of an inconsistent estimator is over-confident, and hence cannot indicate the uncertainty of the estimate. Let the counterpart of $P_{i,k|k}$ and $P_{i,k|k-1}$ obtained

by the ICF be, respectively, $P^{\mathrm{ICF}}_{i,k|k}$ and $P^{\mathrm{ICF}}_{i,k|k-1}$. In the case where the agents communicate only once before updating their local estimates, the update steps of the ICF proposed in [9] can be written as

$$P^{\mathrm{ICF}}_{i,k|k} = [\sum_{j\in J_i}\sigma_{ij}\cdot(P^{\mathrm{ICF}}_{j,k|k-1})^{-1}+N\sum_{j\in J_i}\sigma_{ij}H_j^\top R_j^{-1}H_j]^{-1}, \quad (14)$$

where σ_{ij}, $\forall j$ are the weights related to the rate parameter ε. Therefore, it is very possible that

$$N\sum_{j\in J_i}\sigma_{ij}H_j^\top R_j^{-1}H_j \succeq \sum_{j\in J_i}H_j^\top R_j^{-1}H_j. \quad (15)$$

Note that the right-hand side of the above equation is the total information contained in $\{z_{j,k}\}_{j\in J_i}$. Therefore, it is possible that the $P^{\mathrm{ICF}}_{i,k|k}$ obtained from (14) is smaller than the true error covariance of the local posterior estimates, especially when $N \gg |J_i|$, which is usually the case in sparse wireless sensor networks.

D. Confidence of Estimates

While the consistency is guaranteed, one would also come up with an estimate that is as confident as possible. In the case where the agents communicate only once before local updates, the update step of the approximated error covariance obtained by the KLA algorithm [10], denoted as $P^{\mathrm{KLA}}_{i,k|k}$, has the following form:

$$P^{\mathrm{KLA}}_{i,k|k} = \{\sum_{j\in J_i}\sigma_{ij}[(P^{\mathrm{KLA}}_{j,k|k-1})^{-1}+H_j^\top R_j^{-1}H_j]\}^{-1}, \quad (16)$$

where $\sigma_{ij} = \{\Sigma\}_{(i,j)}$ with Σ being a stochastic matrix associated with the communication graph. In general, given the same set of prior estimates and local measurements from the inclusive neighborhood, the local posterior estimate obtained by (6) is more confident than that obtained by (16). This is mainly due to the following two reasons.

(1) The KLA algorithm did not specify the selection of the weights σ_{ij}, $\forall j$. The only constraint formulated in [10] is Σ being a primitive stochastic matrix. When only one communication step is feasible, the weights really matter in minimizing the fused $P^{\mathrm{KLA}}_{i,k|k}$.

(2) More importantly, even if we suppose that the same set of weights, say σ_{ij}, $\forall j \in J_i$, are selected for both fusion processes in (16) and (6), the update steps of the proposed DHIF are still guaranteed to give more confident estimates since $\sum_{j\in J_i}H_j^\top R_j^{-1}H_j \succeq \sum_{j\in J_i}\sigma_{ij}H_j^\top R_j^{-1}H_j$, which in turn, implies that $P^{\mathrm{KLA}}_{i,k|k} \succeq P_{i,k|k}$.

Superficially, (6) might look similar to (16). However, the philosophies behind the proposed DHIF and the KLA are very different. Compared to the proposed DHIF, the update steps of the KLA can also be regarded as a two-step fusion process. The first step is that, for each $i \in \mathscr{V}$, fuse $z_{i,k}$ and $\hat{x}^{\mathrm{KLA}}_{i,k|k-1}$ into an intermediate estimate $\check{x}^{\mathrm{KLA}}_{i,k}$ with $\check{P}^{\mathrm{KLA}}_{i,k}$ being its approximated error covariance. The fusion strategy (2) is adopted as $z_{i,k}$ and $\hat{x}^{\mathrm{KLA}}_{i,k|k-1}$ have uncorrelated errors. The second step is to fuse $\{\check{x}^{\mathrm{KLA}}_{j,k}\}_{j\in J_i}$. Note that the information sources in this set have unknown correlations in general due to the fact that $\{\hat{x}^{\mathrm{KLA}}_{j,k|k-1}\}_{j\in J_i}$ become highly correlated as the fusion process goes on. Therefore, the consensus on the

PDFs, which is equivalent to the CI algorithm in the case therein (see [10] for details), is used to obtain the posterior estimate and approximated error covariance. Note that during the overall fusion process of the KLA algorithm, the fact that all local measurement noise are mutually uncorrelated is not utilized. By fusing the $\hat{x}^{\mathrm{KLA}}_{i,k|k-1}$ and $z_{i,k}$ first, $z_{i,k}$ is treated as correlated with $z_{j,k}, \forall j \in N_i$ with unknown correlations. This makes the final estimate become relatively conservative.

VI. BOUNDEDNESS OF THE ESTIMATE ERROR

In this section, the sufficient condition for each local estimate error to be upper bounded, is formulated. Specifically, we assume the communication topology and the set of naive agents are both fixed. The scenario in which they are time-varying is left for the future work. Due to the similarities in terms of the structures of the DHIF and KLA algorithms, we borrow the framework of the proofs in [10], and add our extensions. The following lemma will be used in the proof of the main result.

Lemma 4: Let D be the stochastic matrix associated with $\mathscr{G}(\mathscr{V},\mathscr{E})$. Let $\mathscr{R}$ and $\mathscr{B}$ be two exclusive sets of vertices such that $\mathscr{R}\cup\mathscr{B} = \mathscr{V}$. Suppose that $\mathscr{G}(\mathscr{R},\mathscr{E}_\mathscr{R})$ is strongly connected, where $\mathscr{E}_\mathscr{R} \triangleq \{(i,j)|(i,j)\in\mathscr{E},i,j\in\mathscr{R}\}$. Also suppose that there exists a directed path from $\mathscr{R}$ to every agent $j \in \mathscr{B}$ but there is no edge from any agent in $\mathscr{B}$ to $\mathscr{R}$. Then there exists a $\bar{k}$ such that $\{D^k\}_{(i,j)} > 0$, $\forall i \in \mathscr{V}$ and $j \in \mathscr{R}$, if $k \geq \bar{k}$. Moreover, the same conclusion holds for $\{\mathfrak{D}_k\}_{(i,j)}$, where $\mathfrak{D}_k$ is the multiplication of k matrices which belong to the stochastic matrices family induced by D.

Proof: Without loss of generality, let $\mathscr{R} = \{1,\cdots,|\mathscr{R}|\}$ and $\mathscr{B} = \{|\mathscr{R}|+1,\cdots,N\}$. Let $D_\mathscr{R} \in \mathbb{R}^{|\mathscr{R}|\times|\mathscr{R}|}$ be the stochastic matrix associated with the strongly connected graph $\mathscr{G}(\mathscr{R},\mathscr{E}_\mathscr{R})$. Thus $D_\mathscr{R}$ is primitive. Because there is no edge from $\mathscr{B}$ to $\mathscr{R}$, D can be written as a block lower triangular matrix

$$D = \left[\begin{array}{cc} D_\mathscr{R} & \mathbf{0} \\ D_{\mathscr{B}\mathscr{R}} & D_\mathscr{B} \end{array}\right],$$

where $D_{\mathscr{B}\mathscr{R}}$ and $D_\mathscr{B}$ are matrices with appropriate dimensions. It follows that

$$D^k = \left[\begin{array}{cc} D_\mathscr{R}^k & \mathbf{0} \\ D'_{\mathscr{B}\mathscr{R}} & D_\mathscr{B}^k \end{array}\right], \text{ where } D'_{\mathscr{B}\mathscr{R}} = \sum_{r=0}^{k-1}D_\mathscr{B}^r D_{\mathscr{B}\mathscr{R}}D_\mathscr{R}^{k-1-r}.$$

As $D_\mathscr{R}$ is primitive, $D_\mathscr{R}^k > \mathbf{0}$ for any sufficiently large k. Thus it remains to show that $D'_{\mathscr{B}\mathscr{R}} > \mathbf{0}$ for any sufficiently large k. This can be shown by induction.

Suppose that when $k = k'$, $D_\mathscr{R}^k > \mathbf{0}$. Also suppose that agent $i_1 \in \mathscr{B}$ has a direct edge from $\mathscr{R}$. Thus, $\{D\}_{(i_1,j\in\mathscr{R})}$ contains at least one positive entry. It follows that

$$\{D^{k'+1}\}_{(i_1,j\in\mathscr{R})} = \{DD^{k'}\}_{(i_1,j\in\mathscr{R})} = \{D\}_{(i_1,:)}\{D^{k'}\}_{(:,j\in\mathscr{R})}$$
$$\geq \{D\}_{(i_1,j\in\mathscr{R})}D_\mathscr{R}^{k'},$$

where the last inequality is due to the fact that $D \geq \mathbf{0}$. As $\{D\}_{(i_1,j\in\mathscr{B})}$ contains at least one positive entry and $D_\mathscr{R}^{k'} > \mathbf{0}$, it follows that $\{D^{k'+1}\}_{(i_1,j\in\mathscr{R})} > \mathbf{0}$. Moreover, it follows that

as $k' \to \infty$, $D_{\mathscr{R}}^{k'} \to \mathbf{1}v_1^\top$, where $v_1 > \mathbf{0}$ is the left eigenvector associated with the simple eigenvalue 1 of $D_{\mathscr{R}}$ satisfying $\|v_1\|_1 = 1$. That is, each entry in $\{D^k\}_{(i_1, j \in \mathscr{R})}$ is uniformly lower bounded above, for any $k \geq k'$. Let $i_2 \in \mathscr{B}$ be a child vertex of i_1. Thus, $\{D\}_{(i_2,i_1)} > 0$. Since $\{D^{k'+1}\}_{(i_1, j \in \mathscr{R})} > \mathbf{0}$ as previously shown, it follows that

$$\{D^{k'+2}\}_{(i_2, j \in \mathscr{R})} = \{DD^{k'+1}\}_{(i_2, j \in \mathscr{R})}$$
$$= \{D\}_{(i_2,:)}\{D^{k'+1}\}_{(:, j \in \mathscr{R})} \geq \{D\}_{(i_2,i_1)}\{D_{\mathscr{R}}^{k'+1}\}_{(i_1, j \in \mathscr{R})} > \mathbf{0}.$$

By the similar approach, it can also be shown that $\{D^k\}_{(i_2, j \in \mathscr{R})} > \mathbf{0}$ for any $k \geq k' + 1$. Therefore, it can be eventually shown that for any agent i that has a directed path from some leader component $\mathscr{R}$, there exists a sufficiently large $\bar{k}$ such that, for any $k \geq \bar{k}$, $\{D^k\}_{(i,j)} > 0$, $\forall j \in \mathscr{R}$.

As each corresponding entry of all matrices, which belong to the stochastic matrix family induced by the same matrix D, has the same type (zero/nonzero), the same conclusion holds for $\{\mathfrak{D}_k\}_{(i,j)}$ ∎

The following lemma is also used for the proof later on.

Lemma 5 ([10]): Let F be a nonsingular matrix. Then for any $Y \succ \mathbf{0}$ and $\check{\Omega} \succeq \mathbf{0}$, there exists a $\check{\beta} \in (0,1]$ such that $(F\Omega^{-1}F^\top + Y)^{-1} \succeq \check{\beta}F^{-\top}\Omega F^{-1}$ for any $\Omega^{-1} \succeq \check{\Omega}^{-1}$.

Theorem 2: Suppose that Assumptions 1, 2 and Eq. (12) are satisfied. Also suppose that the state transition matrix F is nonsingular. Then for each agent $i \in \mathscr{V}$, there exists $\bar{k} \in \mathbb{Z}^+$ and $\bar{P}_i \succ \mathbf{0} \in \mathbb{R}^{n \times n}$, such that $\bar{P}_i \succeq P_{i,k|k}$, $\forall k > \bar{k}$, if there exists at least one strongly connected subgraph $\mathscr{G}(\mathscr{R}, \mathscr{E}_{\mathscr{R}})$ with joint observability, where $\mathscr{R} \subseteq \mathscr{V}$ and $\mathscr{E}_{\mathscr{R}} \triangleq \{(i,j)|(i,j) \in \mathscr{E}, i,j \in \mathscr{R}\}$ that has a directed path from $\mathscr{R}$ to agent i.

Proof: As the proof follows the framework of Theorem 3 in [10], only the part contributed by this paper will be shown in detail. As shown in Theorem 1, if (12) is satisfied, then for any $k \in \mathbb{Z}^+$, $P_{i,k|k} \succeq \tilde{P}_{i,k|k} \succ \mathbf{0}$. Therefore by Lemma 5, $(FP_{i,k|k}F^\top + BQB^\top)^{-1} \succeq \check{\beta}F^{-\top}P_{i,k|k}^{-1}F^{-1}$. Define $D_k \in \mathbb{R}^{n \times n}$ such that $\{D_k\}_{(i,j)} \triangleq d_{ij}^{(k)}$, where $d_{ij}^{(k)} = \omega_{ij}^{(k)}$ if $j \in J_i$ and $d_{ij}^{(k)} = 0$ if $j \notin J_i$, $\forall i, j$. It follows from (8) that

$$P_{i,k|k}^{-1} = \bar{S}_{i,k} + \sum_j d_{ij}^{(k)}(FP_{j,k-1|k-1}F^\top + BQB^\top)^{-1}$$
$$\succeq \bar{S}_{i,k} + \sum_j d_{ij}^{(k)}\check{\beta}F^{-\top}P_{j,k-1|k-1}^{-1}F^{-1}, \quad (17)$$

Define $\mathfrak{D}_p^q \triangleq D_q D_{q-1} \cdots \bar{D}_p$ for some $q > p$, where $\{\bar{\cdot}\}$ is an element-wised ceil function. Note that $\mathfrak{D}_p^q$ has the same structure with D_k^{q-p+1} for any $k \in \mathbb{Z}^+$. Suppose that $\bar{k} \geq 2$, using the recursive inequality in (17), it follows that $P_{i,k|k}^{-1} \succeq \check{\beta}^{\bar{k}}\sum_j\{\mathfrak{D}_{k-\bar{k}+1}^k\}_{(i,j)}F^{-\bar{k}\top}P_{j,k-\bar{k},k-\bar{k}}F^{-\bar{k}} + \Gamma_i$, where

$$\Gamma_i = \bar{S}_{i,k} + \sum_{\tau=2}^{\bar{k}}\check{\beta}^{\tau-1}\sum_j\{\mathfrak{D}_{k-\tau+2}^k\}_{(i,j)}F^{(1-\tau)\top}H_j^\top R_j^{-1}H_jF^{1-\tau}$$

$$\succeq \sum_{\tau=2}^{\bar{k}}\check{\beta}^{\tau-1}\sum_{j\notin\mathfrak{N}}\{\mathfrak{D}_{k-\tau+2}^k\}_{(i,j)}F^{(1-\tau)\top}H_j^\top R_j^{-1}H_jF^{1-\tau} \triangleq \Gamma_i'.$$

The last inequality is due to the fact that $\bar{S}_{i,k} \succeq \mathbf{0}$ and $R_j^{-1} = \mathbf{0}$, $\forall j \in \mathfrak{N}$. For simplicity, suppose that $\mathscr{G}(\mathscr{R}, \mathscr{E}_{\mathscr{R}})$
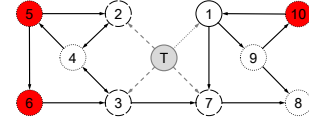

Fig. 1. Communication topology for simulation. T: target; Dashed circle: agents with partial observability; Solid circle: agents with full observability; Dotted circle: agents not observing the target; Red: naive agent.

has a directed path to every other agent[1]. Thus Lemma 4 can be used[2]. Therefore, there exists $\bar{\tau}$ such that if $\tau \geq \bar{\tau}$, $\{\mathfrak{D}_{k-\tau+2}^k\}_{(i,j)} > 0, \forall i \in \mathscr{V}$ and $j \in \mathscr{R}$. Recall that $\check{\beta} > 0$ and $R_j \succ \mathbf{0}$, $\forall j \notin \mathfrak{N}$. It follows that if agents in $\mathscr{R}$ has joint observability, $\Gamma_i' \succ \mathbf{0}$ for any $k \geq \bar{k} \triangleq \bar{\tau} + n$. Let $\bar{P}_i \triangleq (\Gamma_i')^{-1}$. It follows that $\bar{P}_i \succ \mathbf{0}$ and $\bar{P}_i \succeq P_{i,k|k}$. ∎

Theorem 2 formulates the sufficient condition for the bounded local posterior estimate errors. Essentially, the local estimate errors at a certain agent is bounded as long as the agent is able to eventually obtain the knowledge about the entire state from somewhere.

The strategies to decrease the upper bound of $P_{i,k|k}$ for each $i \in \mathscr{V}$ can also be observed from the proof of Theorem 2. Note that $P_{i,k|k}^{-1} \succeq \Gamma_i' \succ \mathbf{0}$ with Γ_i' defined in the proof of sufficiency. This implies that the upper bound of $P_{i,k|k}$ can be decreased if Γ_i' is increased. This is achieved if: 1) there are fewer naive agents so that the summation with respect to j has more positive terms; 2) the graph is denser so that for a certain $\bar{k}$ the summation with respect to τ has more positive terms; 3) local measurements are more accurate (smaller R_j) or redundant (H_j with more rows) so that each term in the summation has a greater value.

VII. SIMULATIONS

In the simulation, the target follows the linear dynamic system in (4). The parameters are adopted from the simulation example used in [10], and listed as follows:

$$F = \begin{bmatrix} \mathbf{I}_2 & \Delta_T\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}, Q = \begin{bmatrix} 5\Delta_T^3\mathbf{I}_2/3 & 5\Delta_T^2\mathbf{I}_2/2 \\ 5\Delta_T^2\mathbf{I}_2/2 & 5\Delta_T\mathbf{I}_2 \end{bmatrix}, B = \mathbf{I}_4,$$

where $\Delta_T = 4$ is the sampling interval. A distributed sensor network with $N = 10$ agents, whose directed communication topology is shown in Figure 1, is used to track the state of the target. Let $H_2 = [1\ 0\ 0\ 0]$, $H_3 = H_7 = [0\ 1\ 0\ 0]$ and $H_1 = [H_2^\top\ H_3^\top]^\top$. Correspondingly, let $R_2 = R_3 = R_7 = 225$, $R_1 = 225\mathbf{I}_2$. Also let $H_i = \mathbf{0}$ and $R_i^{-1} = \mathbf{0}$ for the other agents. Note that there exists no directed spanning tree in this topology. However, the conditions formulated in Theorem 2 are satisfied.

The proposed DHIF algorithm is implemented, where the weights in Step 8 of Algorithm 1 are solved from (7). The KLA algorithm and the ICF are selected for comparisons.

---

[1] When there are multiple disjoint strongly connected subgraphs, the subgraphs can be treated separately.

[2] Note that although the first condition in Theorem 2 does not assume there is no edge path from agent $i \notin \mathscr{R}$ to agent $j \in \mathscr{R}$, this can always be guaranteed by including all $i \notin \mathscr{R}$ which has a directed path to $j \in \mathscr{R}$ into $\mathscr{R}$. Therefore, Lemma 4 can be applied.
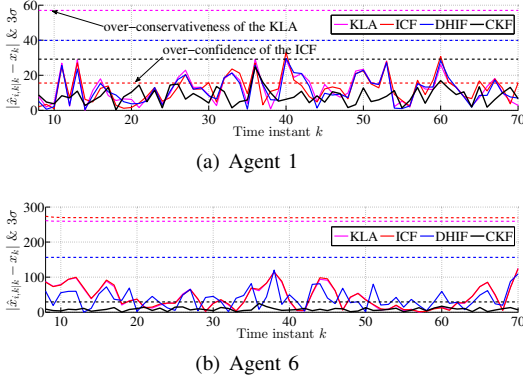
(a) Agent 1

(b) Agent 6

Fig. 2. $|\hat{x}_{i,k|k} - x_k|$ and 3σ-bound given by each algorithm: x-position
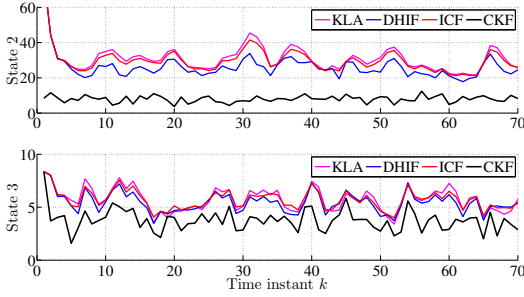

Fig. 3. Comparisons of RMSE $\psi_k = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_{i,k|k}^{(j)} - x_k^{(j)})^2}$ obtained by each algorithm, where the superscript "(j)" is the trial index

As the selection of weights was not specifically mentioned in [10], we let $\sigma_{ij} = 1/|J_i|, \forall i$. The rate parameter ε used in the ICF is selected to be the same as that in [9], i.e., $\varepsilon = 0.65/\Delta_{\max}$. A hypothetical centralized Kalman filter, whose observation matrix has the form of $H \triangleq \text{col}\{H_i\}_{i=1}^{N}$, is used as the benchmark.

The local absolute posterior estimate errors of agent 1 (locally observable) and agent 6 (very naive), obtained by each algorithm, are plotted, respectively, in Figure 2(a) and Figure 2(b). The 3σ-bound corresponding to each algorithm, with σ being the standard deviation (STD) computed from the approximated error covariance by the algorithm, is also plotted in the same color but dashed line. As observed in Figure 2(a), the posterior estimate error of the ICF's exceeds its 3σ-bound frequently. Actually, its 3σ-bound is even less than the one obtained by the CKF, which provides the minimum possible MSE. Therefore, the ICF can be overconfident when the agents communicate with each other for only once before updating their local estimates, as analyzed in Section V-C. In both Figure 2(a) and Figure 2(b), the σ value computed by the KLA is around 1.5 times than that computed by the DHIF.

The same simulation as the previous one are further implemented for 500 trials. The rooted mean squared errors (RMSE) in estimating two state components, denoted as ψ_k, for $k \in [1, 70]$, are plotted in Figure 3. As observed, the proposed DHIF has the lowest RMSE among all distributed algorithms.

VIII. CONCLUSIONS

The problem of distributed state estimation of a linear dynamic process in the wireless sensor networks has been considered. A distributed hybrid information fusion (DHIF) algorithm has been proposed. In the proposed algorithm, each agent communicates with its neighbors for only once, before updating its local estimates and approximated error covariance. In the proposed algorithm, the consistency of estimate is guaranteed to be preserved. Meanwhile, the confidence of each local estimate is improved by efficiently utilizing all information sources available to the local agent. The proposed algorithm is fully distributed and robust against the presence of naive agents. The sufficient condition has been formulated to guarantee that the local estimate error being bounded at steady state.

REFERENCES

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
[2] S. Julier and J. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proceedings of the American Control Conference*, Albuquerque, NM, 1997, pp. 2369–2373.
[3] O. E. Drummond, A. J. Perrella Jr, and S. Waugh, "On target track covariance consistency," in *Signal and Data Processing of Small Targets*, vol. 6236, Kissimmee, FL, 2006, pp. 623 615.1–623 615.14.
[4] W. Ren, R. W. Beard, and D. B. Kingston, "Multi-agent Kalman consensus with relative uncertainty," in *Proceedings of the American Control Conference*, Portland, OR, June 2005, pp. 1865–1870.
[5] M. Alighanbari and J. How, "An unbiased kalman consensus algorithm," in *Proceedings of the American Control Conference*, Minneapolis, MN, June 2006, pp. 3519–3524.
[6] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *Proceedings of the IEEE Conference on Decision and Control*, New Orleans, LA, December 2007, pp. 5492–5498.
[7] ——, "Kalman-consensus filter: Optimality, stability, and performance," in *Proceedings of the IEEE Conference on Decision and Control*, Shanghai, China, December 2009, pp. 7036–7042.
[8] A. T. Kamal, C. Ding, B. Song, J. A. Farrell, and A. Roy-Chowdhury, "A generalized kalman consensus filter for wide-area video networks," in *Proceedings of the IEEE Conference on Decision and Control*, Orlando, FL, 2011, pp. 7863–7869.
[9] A. T. Kamal, J. A. Farrell, and A. K. R. Chowdhury, "Information weighted consensus filters and their application in distributed camera networks," *IEEE Transactions on Automatic Control*, vol. 58, no. 12, pp. 3112–3125, 2013.
[10] G. Battistelli and L. Chisci, "Kullback–leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.
[11] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, January 2007.
[12] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[13] W. Niehsen, "Information fusion based on fast covariance intersection filtering," in *Proceedings of the International Conference on Information Fusion*, vol. 2, Annapolis, MD, 2002, pp. 901–904.